

Teresa Santos-Silva · José Trincão  
Ana Luísa Carvalho · Cecília Bonifácio  
Françoise Auchère · Patrícia Raleiras  
Isabel Moura · José J. G. Moura · Maria João Romão

## The first crystal structure of class III superoxide reductase from *Treponema pallidum*

Received: 2 November 2005 / Accepted: 20 March 2006 / Published online: 6 May 2006  
© SBIC 2006

**Abstract** Superoxide reductase (SOR) is a metalloprotein containing a non-heme iron centre, responsible for the scavenging of superoxide radicals in the cell. The crystal structure of *Treponema pallidum* (Tp) SOR was determined using soft X-rays and synchrotron radiation. Crystals of the oxidized form were obtained using poly(ethylene glycol) and MgCl<sub>2</sub> and diffracted beyond 1.55 Å resolution. The overall architecture is very similar to that of other known SORs but TpSOR contains an N-terminal domain in which the desulforedoxin-type Fe centre, found in other SORs, is absent. This domain conserves the β-barrel topology with an overall arrangement very similar to that of other SOR proteins where the centre is present. The absence of the iron ion and its ligands, however, causes a decrease in the cohesion of the domain and some disorder is observed, particularly in the region where the metal would be harboured. The C-terminal domain exhibits the characteristic immunoglobulin-like fold and harbours the Fe(His)<sub>4</sub>(Cys) active site. The five ligands of the iron centre are well conserved despite some disorder observed for one of the four molecules in the asymmetric unit. The participation of a glutamate as the sixth ligand of some of the iron centres in *Pyrococcus furiosus* SOR was not observed in TpSOR. A possible explanation is that either X-ray photoreduction occurred or there was a mixture of redox states at the start of data collection. In agree-

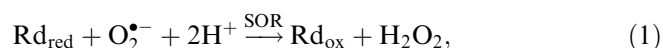
ment with earlier proposals, details in the TpSOR structure also suggest that Lys49 might be involved in attraction of superoxide to the active site.

**Keywords** Superoxide reductase · *Treponema pallidum* · Oxidative stress · Iron–protein · Syphilis

**Abbreviations** Db: *Desulfoarcus baarsii* · Dd: *Desulfovibrio desulfuricans* 27774 · Dfx: Desulfoferrodoxin · Dg: *Desulfovibrio gigas* · Dv: *Desulfovibrio vulgaris* · Dx: Desulforedoxin · FOM: Figure of merit · ICP-OES: Inductively coupled plasma–optical emission spectroscopy · MAD: Multiple-wavelength anomalous diffraction · Pf: *Pyrococcus furiosus* · rmsd: Root mean square deviation · SAD: Single-wavelength anomalous diffraction · SOD: Superoxide dismutase · SOR: Superoxide reductase · Tp: *Treponema pallidum*

### Introduction

Superoxide reductases (SORs) are small non-heme iron containing proteins that catalyse the one-electron reduction of superoxide to hydrogen peroxide, having rubredoxin as their putative electron donor [1–7], according to Eq. 1:



where Rd is rubredoxin. These enzymes participate in oxygen detoxification through a mechanism different from the well-known dismutation of superoxide ion by superoxide dismutase (SOD). They are found only in anaerobic or microaerophilic organisms belonging to Bacteria or Archaea kingdoms and, unlike SOD, only iron has been found at their catalytic site. SOR was first isolated from sulphate-reducing bacteria [8] and its gene was later identified in pathogenic bacteria, such as the syphilis spirochete *Treponema pallidum* (Tp) [3]. SOR was also isolated and characterized from other organ-

This work is dedicated to the memory of Prof. Frank Rusnak. Coordinates and observed structure factor amplitudes have been deposited in the Protein Data Bank under the accession code 1Y07.

T. Santos-Silva · J. Trincão · A. L. Carvalho · C. Bonifácio  
F. Auchère · P. Raleiras · I. Moura · J. J. G. Moura  
M. J. Romão (✉)  
REQUIMTE/CQFB, Departamento de Química,  
Faculdade de Ciências e Tecnologia,  
Universidade Nova de Lisboa,  
2829-516 Caparica, Portugal  
E-mail: mromao@dq.fct.unl.pt  
Tel.: +351-21-2948310  
Fax: +351-21-2948385

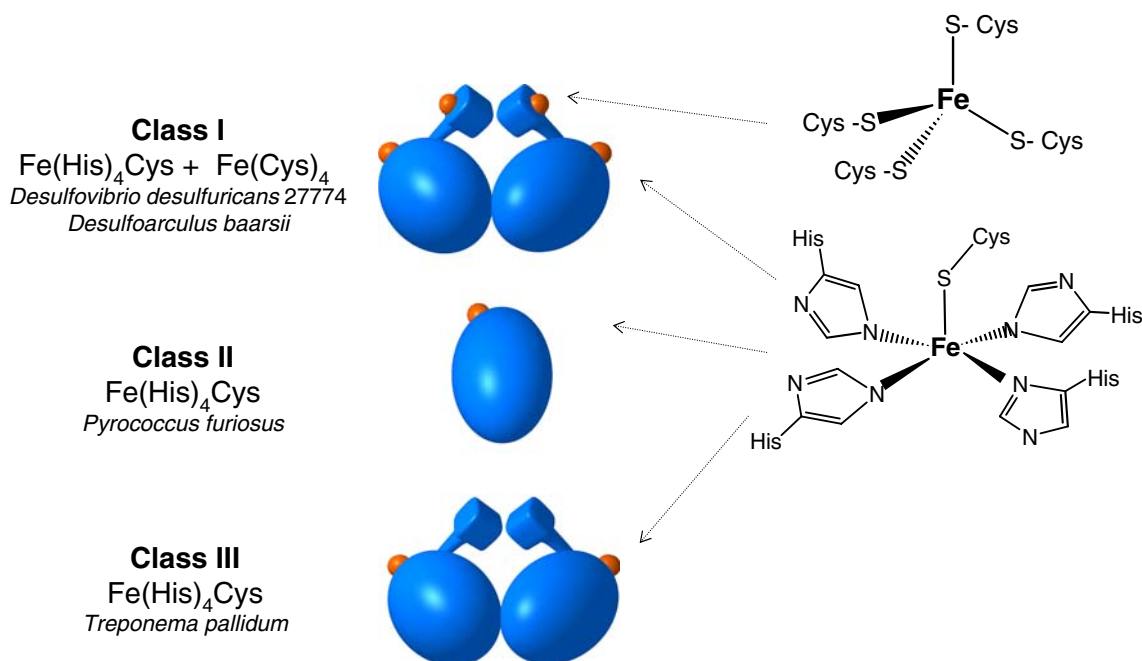
isms such as *Pyrococcus furiosus* (Pf) [4], *Desulfovibrio vulgaris* (Dv) [8] and *Desulfoarculus baarsii* (Db) [5].

SOR can be grouped into three major classes. Members of class I (2Fe-SOR) contain two types of iron centres: a desulfiredoxin (Dx) type centre  $\text{Fe}(\text{Cys})_4$ , and a  $\text{Fe}(\text{His})_4(\text{Cys})$  centre. Members of classes II and III (1Fe-SOR) contain only the  $\text{Fe}(\text{His})_4(\text{Cys})$  catalytic site. Class II is the smallest member of the family and folds into one single domain (Fig. 1). SOR belonging to classes I and III present an additional N-terminal domain, which, in the case of class I, also contains the extra  $\text{Fe}(\text{Cys})_4$  centre. This domain is very similar to the one present in Dx [9, 10]. Classes II and III are also designated as 1Fe short-chain and 1Fe long-chain neelaredoxins, while the 2Fe-SORs are also known by the trivial designation of desulfoferrodoxin (Dfx) [8]. Class I and III SORs are homodimers in solution. The dimerization is stabilized by a net of intersubunit hydrogen bonds in a pattern which is conserved in the two classes of the family.

The iron atom of the active site has an unusual arrangement of four histidines in the equatorial plane and one cysteine in the axial position [11, 12] and only reacts with superoxide in the ferrous state. After one catalytic cycle (reduction of superoxide to hydrogen peroxide), the iron ion is in the ferric state and a further reductive step is required in order to regenerate the active state. An additional ligand has been proposed to be implicated in the catalytic site of SOR. In PfSOR, Glu14 has been assigned as a sixth labile ligand of the iron atom in its ferric state, corresponding to a "resting" state of the enzyme [12]. This proposal was further sustained by Fourier transform IR studies on DbSOR

(Glu47) and TpSOR (Glu48) [13]. Site-directed mutagenesis studies have revealed that this glutamic acid is not rate-limiting in the reaction of SOR with superoxide [14]. Nevertheless, this residue has been implicated in the release of hydrogen peroxide in the proposed two-step mechanism of SOR [15].

Although the sequence of several microbial genomes has revealed the widespread presence of SOR, only some of these proteins have been fully or partially characterized spectroscopically, including SOR from *Tp* [3, 5], *Desulfovibrio gigas* (Dg) [16], *Archaeoglobus fulgidus* [17], *Pf* [18], *Dv*, *Desulfovibrio desulfuricans* (Dd) [5, 8, 19, 20] and *Db* [5]. The gene encoding DvSOR was cloned in two fragments in order to obtain polypeptides corresponding to the N- and C-terminal domains [19]. Structurally, the N-terminal domain mimics Dx, while the C-terminal domain accommodates the  $\text{Fe}(\text{His})_4(\text{Cys})$  catalytic site. These experiments have shown that the two domains of class I SOR (Dfx) preserve the structural properties of the respective centres when expressed separately [19, 21]. This suggests a probable evolutionary gene fusion between the genes for the two domains. In independent work [22] the DvSOR gene was mutated in order to remove the  $\text{Fe}(\text{Cys})_4$  centre, and the resulting gene product was shown to retain the activity of the wild-type enzyme. Although the Dx domain has been proposed to be involved in intramolecular electron transfer towards the catalytic site [19], so far, no experimental evidence has been gathered and its presence in class I SOR (Dfx) remains an intriguing question. The presence/absence of the Dx domain may be a consequence of an evolutionary step from class I SOR to class II SOR.



**Fig. 1** Schematic representation of the three superoxide reductase (SOR) classes. The organism from which a crystal structure has been determined is indicated for each class

Crystal structures of members of classes I and II have been solved: *DdSOR* [11] and *DbSOR* [23] belonging to class I, and *PfSOR* [12] from class II. The structure of *TpSOR* now reported is the first representative of a class III SOR. A more complete view of the versatility of the SOR structures can now be discussed. By comparison and analysis of the amino acid sequences and structural details of the three classes of SOR it is clear that the proteins vary in complexity, but their active site is incorporated in a structurally well-defined site. It is interesting to note that in class III *TpSOR* the N-terminal fold is maintained even in the absence of the so-called structural iron centre.

## Materials and methods

### Crystallization and data collection

*TpSOR* was cloned into vector pTpNeelT77-8/19 and overexpressed in *Escherichia coli* BL21 (DE3) cells. It was purified to homogeneity using a combination of anion exchange and gel filtration chromatography, as described previously [3].

The oxidized form of *TpSOR* was obtained using  $\text{Na}_2\text{IrCl}_6$  as an oxidant. Pure protein was incubated with an excess of  $\text{Na}_2\text{IrCl}_6$ , which was subsequently removed using a desalting column. Crystals were grown using the hanging-drop vapour-diffusion method in the presence of 25% (w/v) poly(ethylene glycol) 3350, 0.2 M magnesium chloride and 0.1 M tris(hydroxymethyl)amino-methane HCl at pH 7.0. Blue plate-shaped crystals of  $0.15 \times 0.05 \times 0.05 \text{ mm}^3$  grew within approximately 12 days. Crystallization details have been described elsewhere [24].

Preliminary X-ray diffraction experiments using monochromatic  $\text{CuK}\alpha$  X-ray radiation from a rotating

anode generator, showed that crystals belong to space group  $C2_2$ , with cell constants  $a = 119.3 \text{ \AA}$ ,  $b = 59.9 \text{ \AA}$ ,  $c = 65.5 \text{ \AA}$  and  $\beta = 104.9^\circ$ . The calculated Matthews coefficient [25],  $2.2 \text{ \AA}^3 \text{ Da}^{-1}$ , indicates the presence of four molecules in the asymmetric unit, with a solvent content of approximately 43%. A highly redundant data set was collected in-house to 1.9- $\text{\AA}$  resolution. A second data collection experiment was performed at beamline BM14 at the ESRF, in Grenoble, France. The same crystal was used to collect data at two wavelengths corresponding to the iron edge absorption peak (1.739  $\text{\AA}$ ) and to a high-energy remote wavelength (1.033  $\text{\AA}$ ). With synchrotron radiation the crystals diffracted beyond 1.55  $\text{\AA}$  resolution. Data collection statistics are summarized in Table 1.

### Structure determination

In spite of the high amino acid sequence homology among all members of this family (Fig. 5), molecular replacement attempts systematically failed. Therefore, single-wavelength anomalous diffraction (SAD) and multiple-wavelength anomalous diffraction (MAD) experiments were performed in order to solve the phase problem.

The in-house collected data set, with an overall completeness of 94.2% and an overall redundancy of 19.2, allowed the calculation of preliminary phases [26, 27]. Although the SAD data were collected far from the iron absorption edge, the corresponding phases revealed the position of the four iron atoms in the asymmetric unit. However, the phases obtained were not of sufficient quality to produce an interpretable electron density map and to allow model building. These preliminary phases had a figure of merit (FOM) of 0.38 and a phasing power of 0.47.

**Table 1** Data collection statistics for *Treponema pallidum* superoxide reductase (*TpSOR*) crystals in the oxidized form

Crystal			
Space group	$C2_2$		
Unit cell ( $\text{\AA}$ )			
$a$	119.3		
$b$	60.0		
$c$	65.6		
$\beta$ ( $^\circ$ )	104.9		
	High-energy remote	Iron absorption peak	In-house $\text{CuK}\alpha$
Data collection statistics			
Wavelength ( $\text{\AA}$ )	1.033	1.739	1.542
No. of observed reflections	366,944	219,096	581,439
No. of unique reflections	53,059	28,909	30,190
Resolution limits ( $\text{\AA}$ )	24.7–1.55 (1.63–1.55)	24.0–2.00 (2.11–2.0)	22.0–1.96 (2.07–1.96)
Redundancy	6.9 (5.4)	7.6 (7.6)	19.2 (18.1)
Completeness (%)	82.4 (82.4)	95.0 (95.0)	94.2 (94.2)
Anomalous completeness (%)		95.0 (92.3)	94.0 (86.9)
$\langle I/\sigma(I) \rangle$	8.8 (4.3)	9.6 (5.3)	8.8 (1.9)
$R_{\text{sym}}$ (%)	4.3 (17.7)	4.5 (11.4)	8.2 (39.6)
$R_{\text{ano}}$ (%)		3.0 (5.6)	2.7 (10.4)

Values in parentheses correspond to the highest resolution shell

$R_{\text{sym}} = \sum |I_i(j) - \langle I(j) \rangle| / \sum I_i(j)$ , where  $I_i(j)$  is the  $i$ th measurement of reflection  $j$  and  $\langle I(j) \rangle$  is the overall weighted mean of  $j$  measurements

The synchrotron data sets provided higher resolution data and a much stronger anomalous signal. The MAD experiment confirmed the four iron sites found by SAD and supplied better phases for model building and refinement. Patterson anomalous difference maps calculated for the two wavelengths showed clear peaks for each iron centre. A combination of density modification protocols and NCS averaging yielded a highly detailed experimental map (FOM of 0.67), where the position of the four molecules was clearly established. At this stage, the *DdSQR* model (PDB accession code 1DFX) [28] was used as a starting model.

### Model building and refinement

Manual model building was straightforward using the experimental map. Apart from a few flexible loops at the surface of the protein, where electron density was uninterpretable, all four polypeptide chains were built and most side chains assigned.

In the first cycles of refinement, data up to 2 Å resolution were used together with NCS restraints for the four molecules of the asymmetric unit. Resolution cutoff was gradually increased and TLS refinement was applied to the two domains of each monomer. The experimental phases, used throughout the refinement, were excluded in the last stages and the NCS copies were refined with no restraints with data up to 1.55 Å resolution.

The refinement cycles resulted in better phases and clearer electron density maps, allowing the localization of some missing side chains at the surface of the protein. However, the electron density for the most exposed loops could not be improved and the amino acid residues in those regions (approximately 5% of the total asymmetric unit) were not included in the final model. Residues Lys11–Gly15 are located at the surface of the protein, and show no continuous electron density in molecules A and B; however, in molecules C and D, the corresponding residues show continuous electron density and model building of this loop was possible. The same was observed for loop Thr42–Ala45, where the electron density is absent for molecules B and C. In molecule A all the residues from this region could be assigned, while in molecule D the electron density for residues Ala45–Glu48 is missing. In this stretch of residues, the refined temperature factors of the side chain atoms are about twice the average temperature factors of the main chain atoms. Water molecules were added using ARP-waters [29] and manually checked with graphics software [30].

After the last round of refinement, the *R*-work and *R*-free factors converged to 18.2 and 23.5%, respectively. The final model has 92% of the residues in the most favoured regions of the Ramachandran plot, with 8% in the additionally allowed regions. Statistical data of the refined model are summarized in Table 2.

### Software

The in-house data set was processed with MOSFLM [31] and SCALA [32], from the CCP4 suite [33], while the HKL2000 package [34] was used for the MAD data sets. Molecular replacement attempts were performed using several programs, such as MOLREP [35], AMoRe [36] and Phaser [37]. The determination of the protein substructure was accomplished with RANTAN [38] implemented in autoSHARP [39], using the in-house data. Later, the HySS program [40], a module of the Phenix package [41], was used for the synchrotron data. Density modification protocols were done with Pirate from the CCP4 suite of programs. Refmac5 [42] was used for refinement and TURBO-FRODO [30] for model building. The geometrical validation was performed with PROCHECK [43]. Sequence alignment analysis was done using ClustalW [44]. The superposition calculations were performed with TURBO-FRODO [30] and the figures were prepared with PyMOL [45] and Alscript [46].

### Inductively coupled plasma–optical emission spectroscopy analysis

In order to investigate the presence of different metal atoms in the oxidized form of *TpSQR*, inductively coupled plasma–optical emission spectroscopy (ICP–OES) was performed. Several crystals grown with the same crystallization conditions were transferred from

**Table 2** Refinement statistics for the *TpSQR* model

Refinement statistics	
Resolution limits (Å)	24–1.55
<i>R</i> factor (%)	18.2
No. of reflections	53,560
<i>R</i> free (%)	23.5
No. of reflections	2,678
No. residues per molecule	128
No. residues per asymmetric unit	489
No. atoms per asymmetric unit	4,091
No. residues missing	22
rmsd bond length (Å)	0.02
rmsd bond angles (Å)	1.765
Average temperature factor (Å <sup>2</sup> )	
Molecule A main chain atoms	11.95
Molecule A side chain atoms	15.27
Molecule B main chain atoms	18.40
Molecule B side chain atoms	20.37
Molecule C main chain atoms	11.31
Molecule C side chain atoms	14.39
Molecule D main chain atoms	16.09
Molecule D side chain atoms	18.76
Water molecules	38.59
Ramachandran plot (%)	
Residues in most favoured regions	92
Residues in additionally allowed regions	8
Residues in disallowed regions	0
Overall <i>G</i> factor	−0.11



the original drops and dissolved in water. The analysis was done with Jobin-Yvon equipment, ULTIMA model, for iron, calcium, magnesium and zinc separately. Significant amounts of iron and magnesium were found; however, neither zinc nor calcium were detected in the sample.

## Results

### Overall structure

The structure solution of *TpSOR* in the oxidized form was achieved by a combination of SAD and MAD techniques, using the anomalous signal of the four iron atoms present in the asymmetric unit, to a resolution of 1.55 Å. The structure revealed four copies (chains A, B, C and D) of the same polypeptide chain in the asymmetric unit of the crystal (Fig. 2a). The four molecules are related to each other by noncrystallographic symmetry operations. Molecule A is related to molecule B by a twofold NCS axis, while molecule C is related to molecule D by a second twofold NCS axis, perpendicular to the first. A third twofold NCS axis emerges, 45° from the other two, roughly parallel to the *b*-axis of the unit cell, relating group AB with group CD. *TpSOR* crystallizes as a dimer of homodimers (AB and CD). Chain A of the *TpSOR* model consists of 121 amino acid residues from a total of 128 that comprise the recombinant protein. The final model includes 367 water molecules. Statistics for the diffraction data processing and model refinement are presented in Tables 1 and 2.

The *TpSOR* monomer is composed of two distinct domains (Figs. 1, 2b). The N-terminal domain (Met1–Arg36) exhibits a Dx-like fold with four  $\beta$ -strands, similar to the N-terminal domain of class I SOR. The C-terminal domain (Ala45–Lys125), harbours the catalytic Fe(His)<sub>4</sub>(Cys) site and displays an immunoglobulin-like fold, with seven  $\beta$ -strands and a short  $\alpha$ -helix, in analogy to the C-terminal domain of classes I and II SOR. As mentioned before *TpSOR* is a homodimer in solution, with an overall arrangement resembling that of classes I and II dimers (Fig. 2b). At the N-terminal domain, dimerization produces two four-stranded antiparallel  $\beta$ -sheets (Fig. 3a), similar to the *DgDx* structure [9]. At the C-terminal domain an open  $\beta$ -barrel is formed. One side of the barrel is defined by an eight-stranded antiparallel  $\beta$ -sheet showing the typical twist between the first and eighth strands, while the other side is formed by a six-stranded  $\beta$ -sheet.

Between the N- and the C-terminal domains of each dimer, a small cavity is found. In class I *DdSOR* and *DbSOR*, this cavity is approximately 430 Å<sup>3</sup> and holds a calcium ion. This ion is coordinated by the side chains of amino acid residues of the C-terminal domain, and it was considered to be crucial for the stabilization of the dimers in the crystals [11]. In class III *TpSOR* no calcium ion is found at this position. In fact, a much

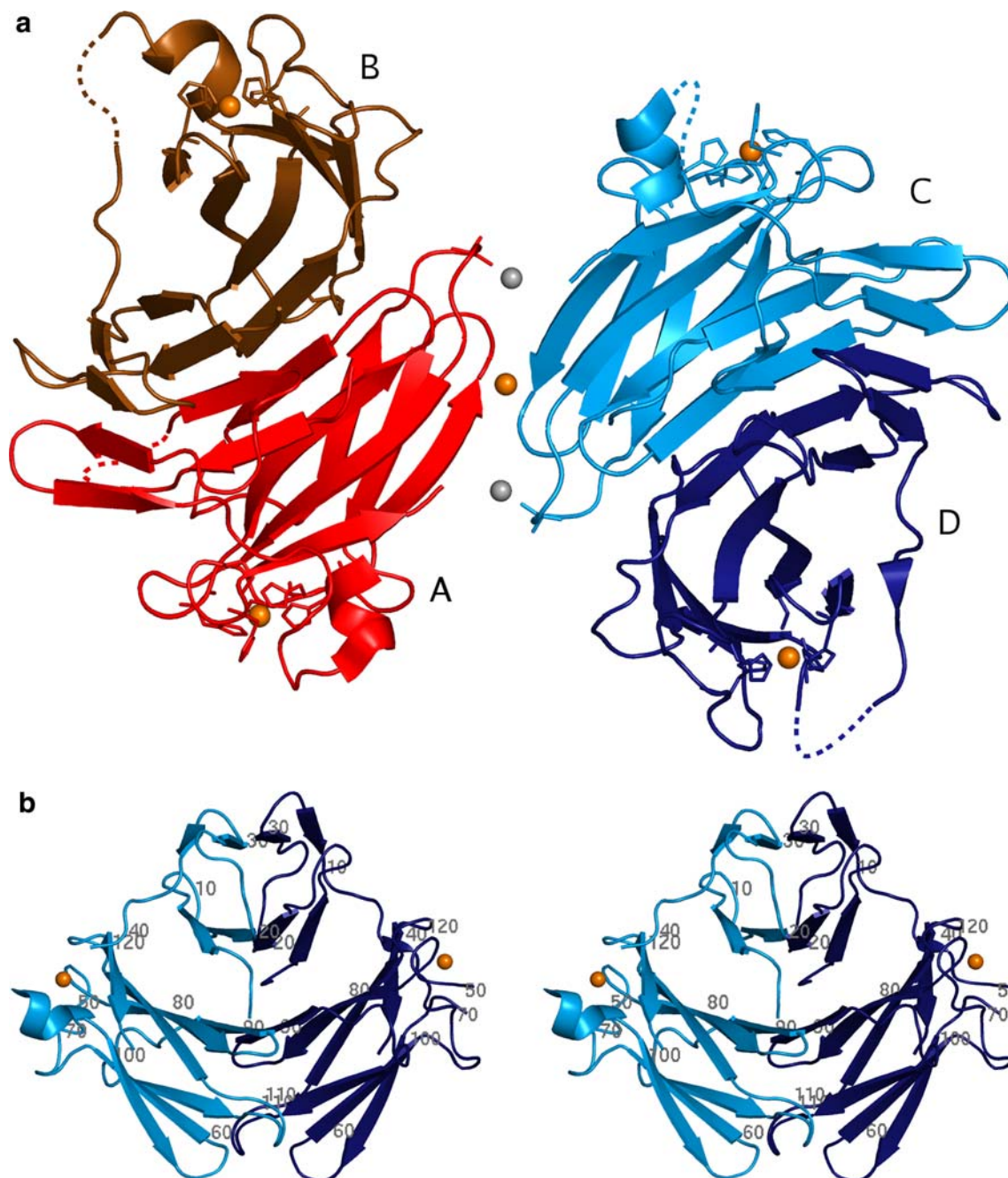
smaller cavity is formed (approximately 180 Å<sup>3</sup>) owing to a tighter packing of the domains. The residues responsible for Ca<sup>2+</sup> coordination in *DdSOR* (Ser87 and Thr89) have been replaced by hydrophobic residues in *TpSOR* (Ile89 and Leu91). Besides making it impossible to coordinate any ions, these residues are important for dimer stabilization. The amino acid sequences of other class III SORs such as *Clostridium acetobutylicum* SOR [47] or *Treponema denticola* SOR [48] also have hydrophobic residues at these positions. These results suggest that Ca<sup>2+</sup> is not needed to maintain the dimer structure in *TpSOR* and will probably also be absent in other class III SOR proteins.

The four molecules in the asymmetric unit are very similar to each other with a root mean square deviation (rmsd) of approximately 0.3 Å for 100 C $\alpha$  atoms (out of 128 amino acids). However, there are differences in three loops where the poor quality of the electron density maps suggests disorder. The first of these loops, Phe8–Phe16, corresponds to the rubredoxin “knuckle” [49] and, together with the second loop (Cys30–Glu33), corresponds to the iron binding motif of Fe(Cys)<sub>4</sub> in *DgDx* and in class I SOR. This centre is absent in *TpSOR*. Three of the four cysteine residues that coordinate the iron atom in *DgDx* and in class I SOR are not present in *TpSOR*, which may explain the increased disorder of the domain. Molecules A and B show no continuous electron density in the experimental electron density maps for the rubredoxin “knuckle”, which is well defined only in molecules C and D. The third loop (Thr42–Glu48) is part of the linker segment between the N- and the C-terminal domains and it could be completely traced only for molecule A. In molecules B and D, the poor quality of the electron density maps impaired complete model building of this segment.

### The active site: iron centre [Fe(N<sup>ε</sup>His)<sub>3</sub>(N<sup>δ</sup>His)(S<sup>γ</sup>–Cys)]

The iron centre is located in the C-terminal domain, close to the linker segment between the N- and the C-terminal domains. It is coordinated by four histidines (N<sup>ε</sup>His50, N<sup>ε</sup>His70, N<sup>ε</sup>His76, N<sup>δ</sup>His122) and one cysteine (S<sup>γ</sup>–Cys119) (Fig. 4) in a square-pyramidal geometry, and is highly exposed to the solvent, similar to what has been described for other SORs. In the homodimer, the two centres lie at opposite ends, 44 Å apart from each other. All four centres have been modelled with full occupancy.

The distances between the iron and the coordinating atoms are in agreement with the other SOR structures (Table 3). The relatively wide range of values for the bond distances can be possibly explained by X-ray photoreduction and/or partial Zn occupancy (see later). The five ligands of the metal have well-defined density in the 2*mF*<sub>o</sub>–*DF*<sub>c</sub> maps. However, in molecule D weak electron density can be observed for the equatorial ligand His76 and its precedent residue Glu75, which exhibit significantly higher temperature factors (28 and



**Fig. 2** Crystal structure of the oxidized form of *Treponema pallidum* (*Tp*) SOR. **a** One asymmetric unit of *Tp*SOR showing the two functional dimers (*AB* and *CD* in two shades of red and blue, respectively) and viewed along one of the local dyads. The four iron atoms, approximately 40 Å apart within each dimer,

are shown as *gold spheres* as well as the extra putative iron site at the interface of molecules *A* and *C*. The two magnesium ions are shown in grey. The disordered loops are represented by *dashed lines*. **b** Stereo ribbon diagram of the *CD* homodimer numbered every tenth residue. The iron atoms are represented as *gold spheres*

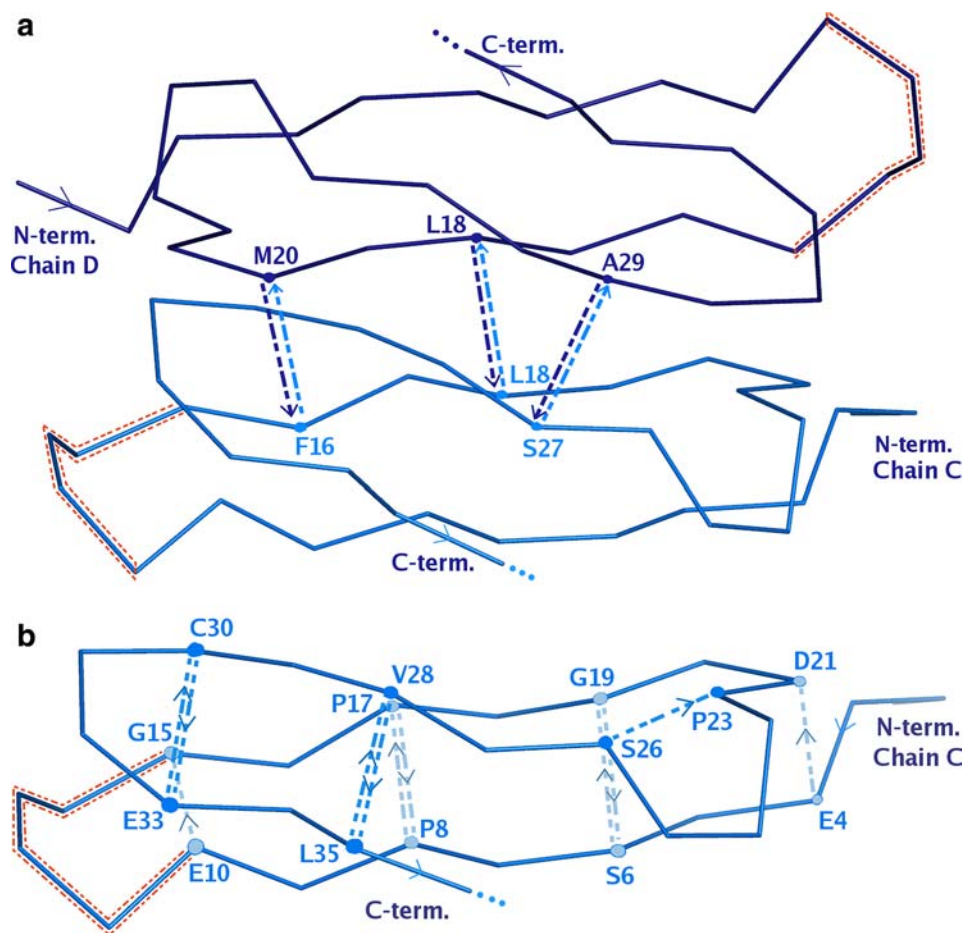
31 Å<sup>2</sup>, respectively) when compared with the other ligands (18 Å<sup>2</sup> on average). This may be due to disorder of the loop 73–77 and/or to crystal packing effects since molecule *D* has fewer crystal contacts in comparison with the other molecules.

The structure of *Pf*SOR in the oxidized form [12] shows, for only half of the molecules in the asymmetric unit, a sixth labile ligand (Glu14). This glutamate

belongs to a disordered loop and is coordinated to the iron atom in the ferric state, with an Fe–O distance of 2.1 Å.

In *Tp*SOR the corresponding loop (Thr42–Glu48) is also disordered but Glu48 (Glu14 in *Pf*SOR) is located 10–12 Å away from the iron centre (where only for molecule *D*, Glu48 could not be traced). Since crystals were grown after treatment of the protein with an oxi-

**Fig. 3** Secondary structure representation of main chain hydrogen bonding pattern (NH  $\rightarrow$  O) at the N-terminal domain. **a** Representation of the hydrogen bonds between subunits C and D. The exposed and highly flexible rubredoxin “knuckle” E10–G15 is shown in *dashed lines* for both monomers. **b** Representation of the main chain hydrogen-bonding pattern, within monomer C



dizing agent, this result was unexpected. As suggested for *Pf*SOR [12], X-ray photoreduction could affect the oxidation state of the *Tp*SOR centre, preventing hexacoordination of the iron atom. In fact, for *Db*SOR [23] some of the iron centres are photoreduced within 3 s of X-ray exposure, although it seems unlikely that a motion of approximately 10 Å occurs at 100 K.

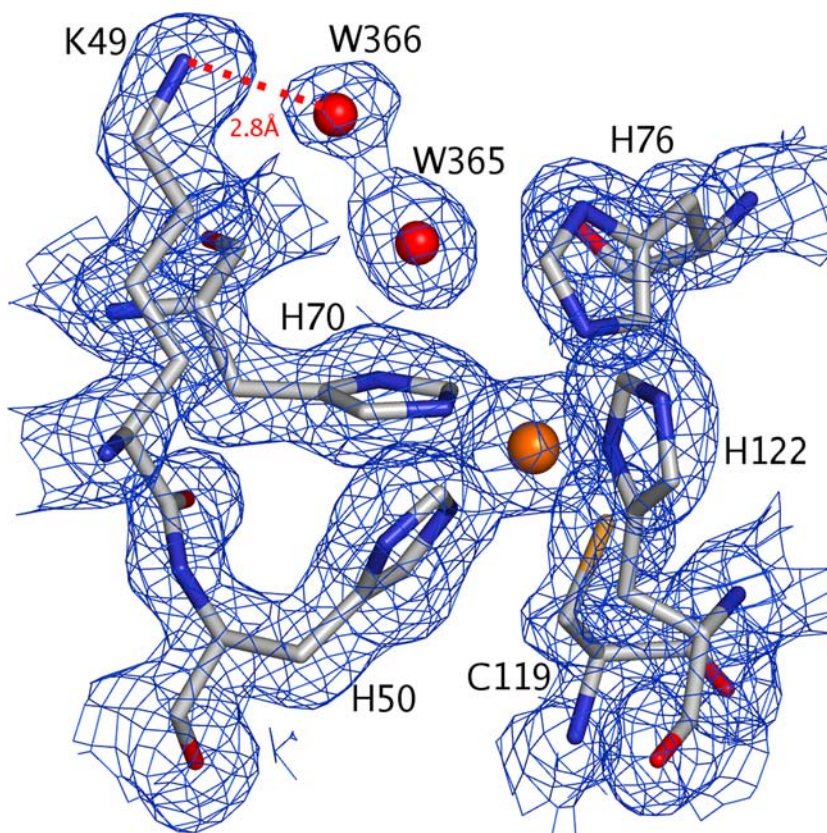
In all SOR structures reported so far, a positive difference electron density is observed at about 4.5 Å from the iron atom, in the free axial position. This has been assigned to the effect of a chloride ion in *Db*SOR, but in *Dd*SOR it could not be identified. In *Tp*SOR a water molecule was modelled in this density, with *B* factors (35–40 Å<sup>2</sup>) comparable to the average *B* factor of waters (38.6 Å<sup>2</sup>) (Table 1). In one of the molecules (C) (Fig. 4) this water molecule contacts with Lys49, via a second water molecule. This is a highly conserved lysine and has been described as an important residue for the catalytic activity of SOR [15], attracting the superoxide ion to the active site. Comparing the H<sub>2</sub>O–Fe distances (Table 3) it is noteworthy that, while for molecules A–C the distances are approximately 4.3 Å, for molecule D the corresponding water molecule (W360) is only 2.7 Å away from the Fe atom. In fact, in *Pf*SOR, a water molecule was also found at approximately 2.6 Å from

the Fe atom, for the two molecules where Glu14 is not bound [12].

Anomalous difference Patterson maps were calculated for the two data sets, collected at two different wavelengths: the iron absorption peak (1.739 Å) and a higher-energy remote wavelength (1.033 Å). As expected, strong anomalous difference peaks, corresponding to the four iron centres in the asymmetric unit, were observed using the structure factors of the iron absorption peak data set. These anomalous difference peaks were also detected in the Patterson maps of the higher-energy data set. Even though the anomalous scattering properties of iron extend over a wide range of wavelengths, we would have expected that the observed anomalous density would be smaller when compared with the peak data. This suggests that a different metal could be competing for the same position. Overexpression of *Tp*SOR in *E. coli* is known to result in a mixture of both iron- and zinc-bound forms of the protein [3]. In fact, at the remote wavelength (1.033 Å), zinc exhibits a higher anomalous signal than iron. Although the anomalous difference intensities for the peak data are similar among the four molecules, the anomalous intensities for the remote data set (1.033 Å) are different, which could be due to the presence of zinc in some of the



**Fig. 4** Details of the active site of molecule C and representation of the final  $2mF_o - DF_c$  electron density map contoured at  $0.8\sigma$  level obtained from the high-energy remote data set. The four histidine residues and the cysteine residue are coordinated to the iron atom. Lys49 is connected to the iron atom through a net of hydrogen bonds mediated by solvent molecules W366 and W365. W365 is at 4.2 Å from the metal atom



**Table 3** Distances between the iron atom and its ligands in the active site [ $\text{Fe}(\text{N}^\epsilon\text{His})_3(\text{N}^\delta\text{His})(\text{S}^\gamma\text{-Cys})$ ]

	Molecule A	Molecule B	Molecule C	Molecule D
Distance Fe–N(His) (Å)				
His50, N $^\epsilon$	2.2	2.3	2.1	2.2
His70, N $^\epsilon$	2.2	2.4	2.2	2.4
His76, N $^\epsilon$	2.4	2.1	2.3	2.2
His122, N $^\delta$	2.3	2.1	2.2	2.2
Distance Fe–S(Cys) (Å)	2.3	2.1	2.3	2.5
Distance Fe–O(W) (Å)	4.3	4.3	4.4	2.7

crystal molecules. Metal atom analysis (ICP–OES) was performed in dissolved crystals; however, the analysis did not reveal the presence of zinc, probably owing to the low amount of the metal present in the crystals. Additional information could be obtained by collecting data above and below the zinc absorption edge, but this could not be performed. Furthermore, ICP analysis did not detect the presence of calcium atoms, in agreement with what was discussed before.

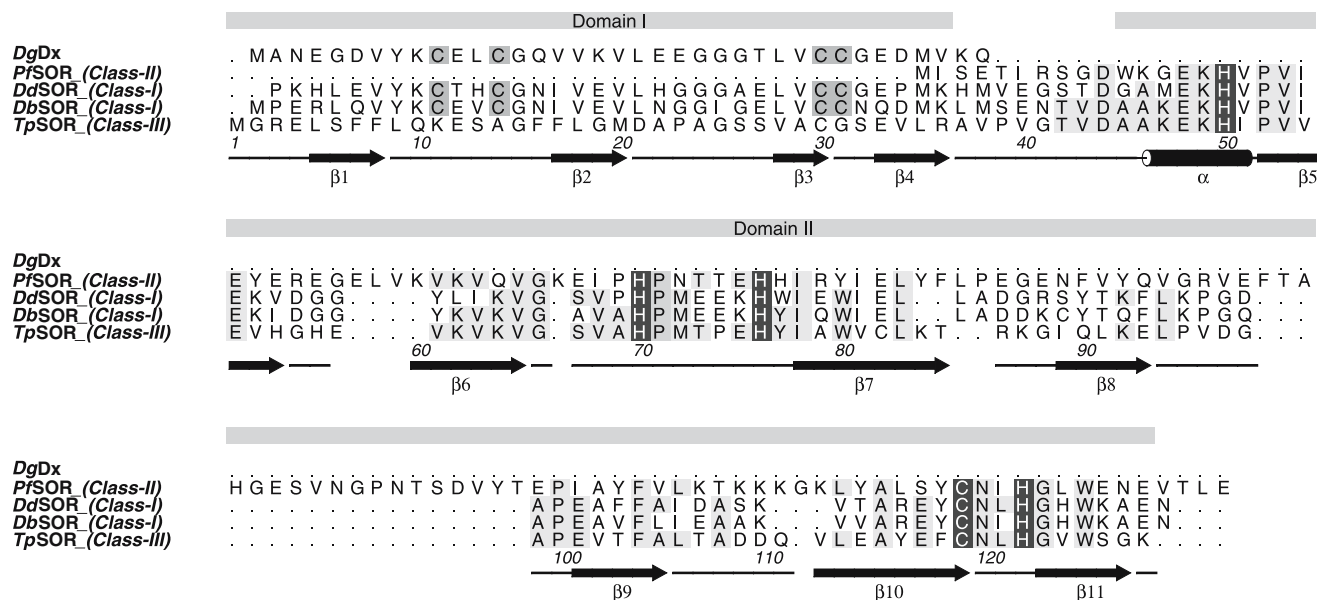
Together with the anomalous signals of the four iron centres, another unexpected strong peak was present in the anomalous difference Fourier map, with intensity similar to that of the iron centres. It is located at the interface between molecules A and C, sitting on the twofold NCS axis that relates the two homodimers and lies approximately 3.3 Å from Gln111A, Gln111C

carboxylic groups and four water molecules. Even though the distances between this strong electron density and potential ligands imply the absence of coordination, it possibly stabilizes the crystallographic dimer (Fig. 2). When included in the model with an occupancy of 0.8, this putative iron atom refined to a  $B$  factor of  $22.4 \text{ Å}^2$ . At this interface two magnesium ions are also present with an octahedral coordination to five water molecules, and to the carbonyl groups of Ala108A and Ala108C. Since  $\text{Mg}^{2+}$  is present in the crystallization conditions (0.2 M  $\text{MgCl}_2$ ) and is bound to the surface of the protein, its presence does not have functional relevance. The two magnesium ions and the putative iron atom are aligned, forming an axis perpendicular to the NCS axis. Together, these two axes form a plane that divides the asymmetric unit into two identical parts.

#### Comparison with related structures

*TpSOR* is structurally very similar to the other SOR proteins. Primary structure alignments show high sequence identity (more than 30%) among the three classes, with amino acid sequence conservation mainly in the C-terminal domain (Fig. 5). The superposition of the *TpSOR* structure with those of the related proteins, represented in Fig. 6, reveals a similar overall architecture for the three classes, in particular for the active site domain. The CD homodimer was chosen for the





**Fig. 5** Amino acid sequence alignment of *Desulfovibrio gigas* desulfuroredoxin (DgDx) and SORs from *Pyrococcus furiosus* (PfSOR), *Desulfovibrio desulfuricans* (DdSOR), *Desulfoarculus baarsii* (DbSOR) and *TpSOR*. Identity to *TpSOR* is indicated with light grey boxes. Residue numbers refer to the sequence of *TpSOR*.

superpositions since the AB homodimer has several disordered regions.

The structural superposition of *TpSOR* and *PfSOR* was carried out using only the C-terminal domain. The low rmsd value found for 80% of the C $\alpha$  atoms (0.8 Å) denotes the high similarity between the two proteins. The main difference is in the length of the loop regions, which are longer in *PfSOR* when compared with the other SOR structures. Structural comparison of *TpSOR* and *DbSOR* reveals more differences in both domains. The high structural similarity observed in the C-terminal domains of the two proteins is not present in the N-terminal domains (Fig. 6b).

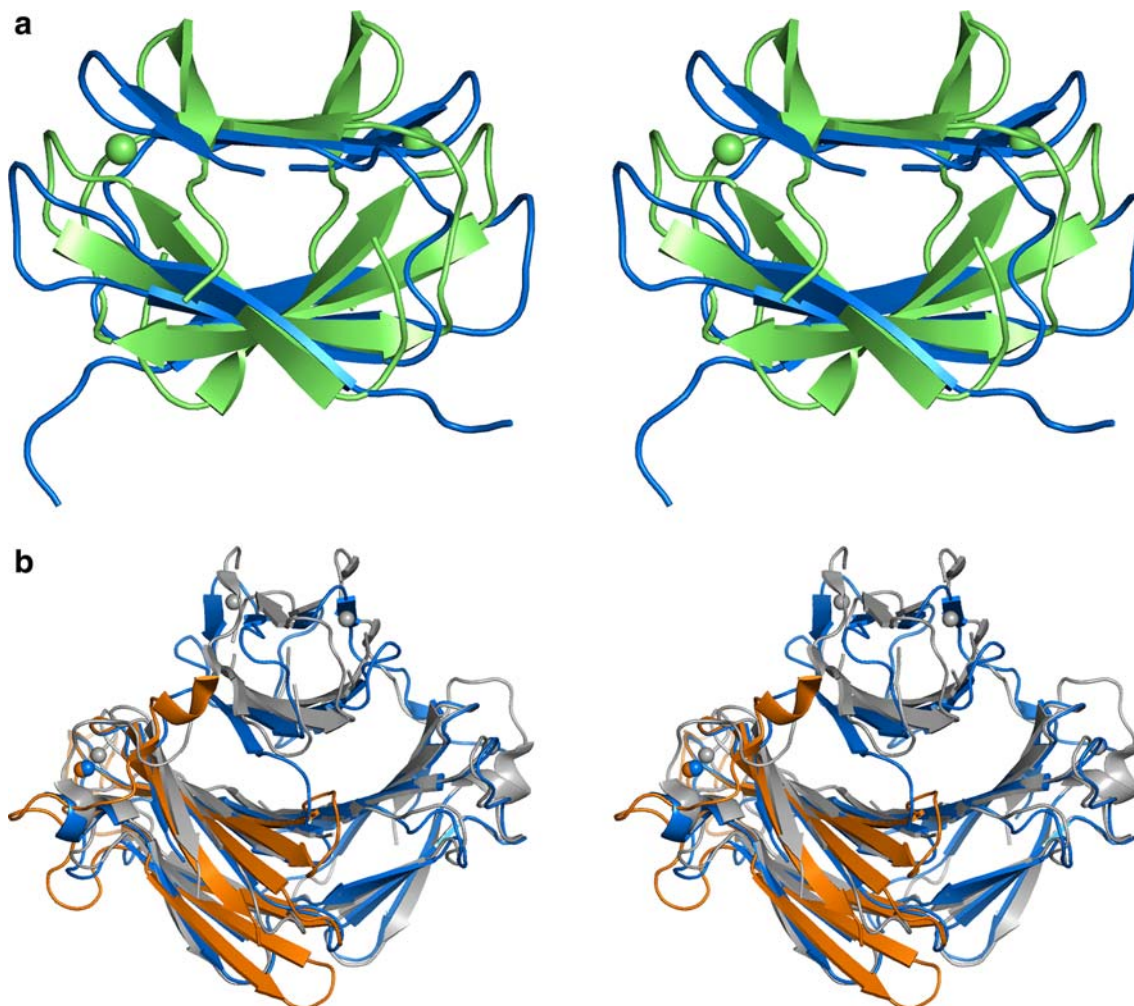
The low rmsd value found for the C-terminal domain (0.7 Å for the superposition of the monomers), doubles for the N-terminal domain, for approximately 90% of the C $\alpha$  atoms (1.7 Å for the same superposition). The amino acid sequence homology between the *TpSOR* N-terminal domain and the same domain of class I SOR or DgDx is not significant, suggesting structural divergence. As mentioned before, the iron binding motifs of centre I Fe(Cys) $_4$  are absent in *TpSOR*; therefore, the characteristic NH  $\rightarrow$  S hydrogen bonding pattern described for the rubredoxin and Dx iron sites [9, 10, 50] does not exist and disorder is observed in the corresponding loops, especially for the AB dimer. In molecules C and D it is clear that the  $\beta$ -strands forming the domain are in different positions and with a different orientation than in DgDx. The two loops that enclose the Fe(Cys) $_4$  centre in DgDx, are approximately 1.5 Å closer to each other in *TpSOR*, where the metal centre is absent. In fact, for DgDx as well as for rubredoxin [9, 10, 50] a structural role was assigned to the Fe(Cys) $_4$

centre, which was therefore considered relevant for the general stability of the structure.

However, the lack of the iron centre and its ligands justifies the larger mobility of the N-terminal domain in *TpSOR*, which is held together by  $\beta$ -sheet interactions. In addition, in *TpSOR* there are fewer intersubunit hydrogen bonds stabilizing the dimer (Fig. 3a) as well as the  $\beta$  structure of each monomer (Fig. 3b). From ten intersubunit NH  $\rightarrow$  O hydrogen bonds found in DgDx [9], only six are conserved in *TpSOR* (Fig. 3a), while in each monomer only 11 hydrogen bonds between adjacent strands are found, instead of the 14 in Dx.

## Discussion and conclusions

*TpSOR* is the first member of class III of the SOR family to be structurally characterized. The high resolution structure now reported provides valuable information for the understanding of this class of enzymes. The effect of the absence of an iron centre Fe(Cys) $_4$  at the N-terminal domain on the protein fold could not be anticipated in the absence of the crystal structure, when compared with class I SOR (*DdSOR* and *DbSOR*). Although the overall arrangement of the protein is very similar to that of the other classes of SOR, the stability of the N-terminal domain (a small  $\beta$ -barrel) is affected by the absence of the iron ion and its ligands. In the related Dx protein, a small homodimeric (2  $\times$  4 kDa) protein comprising only this domain, a clear structural role was assigned to the Fe(Cys) $_4$  centre [9]. The stability of the  $\beta$ -barrel was explained on the basis of a network of intersubunit and intrasubunit hydrogen bonds, as well



**Fig. 6** Stereo view of the ribbon representation of the superpositions of the three SOR classes and DgDx. **a** The N-terminal domain of *TpSOR* (blue) and DgDx (green) (root mean square deviation, *rmsd*, of 2.2 Å for 60 C $\alpha$  atoms out of 72). **b** The CD homodimer

from *TpSOR* (blue) with *DbSOR* (grey) (*rmsd* of 1.7 Å for 209 C $\alpha$  atoms out of 256) and with one monomer of *PfSOR* (orange) (*rmsd* of 0.8 Å for 64 C $\alpha$  atoms out of 80)

as to a characteristic pattern of NH  $\rightarrow$  S interactions between the iron atom and its cysteine ligands. Similar noncovalent interactions are conserved within the N-terminal domain of class I SOR (*DbSOR* and *DdSOR*). In *TpSOR*, in spite of the conserved global fold, the quality of the model reveals some disorder in the corresponding domain, particularly in the region that would harbour the iron centre. This can be explained by the lack of several of the mentioned hydrogen bonds. In addition, since three of the four coordinating cysteines are absent in *TpSOR* the pattern of NH  $\rightarrow$  S bonds is not present.

The interactions within the functional dimers observed for class I SOR structures are also observed in *TpSOR*. The relevance of the calcium ion upon dimerization in *TpSOR* is, however, different from what has been proposed for class I SOR. The hydrophobic residues that replace the calcium ligands present in classes I and II SOR are conserved in other proteins of class III;

hence, on the basis of the *TpSOR* structure, we suggest that calcium is probably not essential for the formation of the functional dimer in class III SOR.

The active site is highly conserved in the three classes of SOR and the iron atom is fivefold coordinated. In the oxidized form of *TpSOR*, the participation of Glu48 as a sixth ligand of the iron centre was expected, as observed in *PfSOR*. However, Glu48 is placed 10–12 Å away from the metal and belongs to a disordered loop. This residue is highly conserved in SOR and has been implicated in the mechanism of release of hydrogen peroxide [15]. It is known that X-ray photoreduction affects the iron centre (*PfSOR* [12] and *DbSOR* [23]) and in *TpSOR* this could be a possible explanation for the presence of a pentacoordinated iron atom. However, a motion of approximately 10 Å seems unlikely to occur at 100 K and further experiments are needed to clarify this point.

The interaction of Lys49 with the Fe atom is mediated by two water molecules, substantiating the

hypothesis that this residue may be involved in the enzymatic mechanism, attracting the superoxide ion into the active site [15]. At the Thr73–Tyr77 loop region, which includes one of the Fe ligands (His76), some disorder is observed for one of the SOR molecules (molecule D). Whether this is due to packing effects or is mechanistically relevant cannot be established at this stage. Furthermore, the existence of a well-ordered water molecule 2.7 Å from the iron atom, also in molecule D, could mean a stabilization of the centre at an intermediate state of the reaction cycle of SOR. These hypotheses need, however, to be investigated through additional structural studies, in particular with reduced and substrate analogues or inhibitor-bound forms of *Tp*SOR.

**Acknowledgements** This work was supported in part by Fundação para a Ciência e Tecnologia, SFRH/BD/6358/2001 (T.S.-S.), BPD-9444/2002 (J.T.) and COST D21 – WG6 (I.M., J.J.G.M. and M.J.R.). The authors would like to thank Mário Sadio for his help in the preparation of Fig. 1, Hassan Berhali, from beamline BM14, for his help during synchrotron data collection at the ESRF, Grenoble, France, and the EMBL EU-HCMP programme for supporting travel and accommodation expenses. The authors would also like to thank the Analytical services of laboratory of REQUIMTE for the ICP-OES results.

## References

- Emerson JP, Coulter ED, Phillips RS, Kurtz DM Jr (2003) *J Biol Chem* 278:39662–39668
- Lombard M, Touati D, Fontecave M, Niviere V (2000) *J Biol Chem* 275:27021–27026
- Jovanovic T, Ascenso C, Hazlett KR, Sikkink R, Krebs C, Litwiller R, Benson LM, Moura I, Moura JJ, Radolf JD, Huynh BH, Naylor S, Rusnak F (2000) *J Biol Chem* 275:28439–28448
- Jenney FE Jr, Verhagen MF, Cui X, Adams MW (1999) *Science* 286:306–309
- Lombard M, Fontecave M, Touati D, Niviere V (2000) *J Biol Chem* 275:115–121
- Emerson JP, Coulter ED, Cabelli DE, Phillips RS, Kurtz DM Jr (2002) *Biochemistry* 41:4348–4357
- Auchère F, Sikkink R, Cordas C, Raleiras P, Tavares P, Moura I, Moura JJ (2004) *J Biol Inorg Chem* 9:839–849
- Moura I, Tavares P, Moura JJ, Ravi N, Huynh BH, Liu MY, LeGall J (1990) *J Biol Chem* 265:21596–21602
- Archer M, Huber R, Tavares P, Moura I, Moura JJ, Carrondo MA, Sieker LC, LeGall J, Romao MJ (1995) *J Mol Biol* 251:690–702
- Archer M, Carvalho AL, Teixeira S, Moura I, Moura JJ, Rusnak F, Romao MJ (1999) *Protein Sci* 8:1536–1545
- Coelho AV, Matias PM, Fulop V, Thompson AW, Gonzalez A, Carrondo MA (1997) *J Biol Inorg Chem* 2:680–689
- Yeh AP, Hu Y, Jenney FE Jr, Adams MW, Rees DC (2000) *Biochemistry* 39:2499–2508
- Berthomieu C, Dupeyrat F, Fontecave M, Vermeglio A, Niviere V (2002) *Biochemistry* 41:10360–10368
- Lombard M, Houee-Levin C, Touati D, Fontecave M, Niviere V (2001) *Biochemistry* 40:5032–5040
- Nivière V, Asso M, Weill CO, Lombard M, Guigliarelli B, Favaudon V, Houee-Levin C (2004) *Biochemistry* 43:808–818
- Chen L, Sharma P, Le Gall J, Mariano AM, Teixeira M, Xavier AV (1994) *Eur J Biochem* 226:613–618
- Abreu IA, Saraiva LM, Carita J, Huber H, Stetter KO, Cabelli D, Teixeira M (2000) *Mol Microbiol* 38:322–334
- Clay MD, Jenney FE Jr, Hagedoorn PL, George GN, Adams MW, Johnson MK (2002) *J Am Chem Soc* 124:788–805
- Ascenso C, Rusnak F, Cabrito I, Lima MJ, Naylor S, Moura I, Moura JJ (2000) *J Biol Inorg Chem* 5:720–729
- Tavares P, Ravi N, Moura JJ, LeGall J, Huang YH, Crouse BR, Johnson MK, Huynh BH, Moura I (1994) *J Biol Chem* 269:10504–10510
- Rusnak F, Ascenso C, Moura I, Moura JJ (2002) *Methods Enzymol* 349:243–258
- Emerson JP, Cabelli DE, Kurtz DM (2003) *Proc Natl Acad Sci USA* 100:3802–3807
- Adam V, Royant A, Niviere V, Molina-Heredia FP, Bourgeois D (2004) *Structure (Camb)* 12:1729–1740
- Santos-Silva T, Trincao J, Carvalho AL, Bonifacio C, Auchere F, Moura I, Moura JJG, Romao MJ (2005) *Acta Crystallogr F* 61:967–970
- Matthews BW (1968) *J Mol Biol* 33:491–497
- Dauter Z, Dauter M, de La Fortelle E, Bricogne G, Sheldrick GM (1999) *J Mol Biol* 289:83–92
- Weiss MS, Sicker T, Hilgenfeld R (2001) *Structure (Camb)* 9:771–777
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) *Nucleic Acids Res* 28:235–242
- Perrakis A, Morris R, Lamzin VS (1999) *Nat Struct Biol* 6:458–463
- Roussel A, Fontecilla-Camps JC, Cambillau C (1990) XV IUCr Congress Bordeaux, France
- Leslie AGW (1992) *Newslett Protein Crystallogr* 26
- Kabsch W (1988) *J Appl Crystallogr* 21:67–71
- Collaborative Computational Project Number 4 (1994) *Acta Crystallogr D* 50:760–763
- Otwinowski Z, Minor W (1997) *Macromol Crystallogr* 276(Pt A):307–326
- Alexei Vagin AT (1997) *J Appl Crystallogr* 30:1022–1025
- Navaza J (1994) *Acta Crystallogr A* 50:157–163
- Read RJ (2001) *Acta Crystallogr D* 57:1373–1382
- Jia-xing Y (1981) *Acta Crystallogr A* 37:642–644
- delaFortelle E, Bricogne G (1997) *Macromol Crystallogr* 276(Pt A):472–494
- Grosse-Kunstleve RW, Adams PD (2003) *Acta Crystallogr D* 59:1966–1973
- Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC (2002) *Acta Crystallogr D* 58:1948–1954
- Murshudov GN, Vagin AA, Dodson EJ (1997) *Acta Crystallogr D* 53:240–255
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) *J Appl Crystallogr* 26:283–291
- Pearson WR (1998) *J Mol Biol* 276:71–84
- DeLano WL (2002) The PyMOL molecular graphics system. DeLano Scientific, San Carlos
- Barton GJ (1993) *Protein Eng* 6:37–40
- Nolling J, Breton G, Omelchenko MV, Makarova KS, Zeng QD, Gibson R, Lee HM, Dubois J, Qiu DY, Hitti J, Wolf YI, Tatusov RL, Sabathe F, Doucette-Stamm L, Soucaille P, Daly MJ, Bennett GN, Koonin EV, Smith DR, Finishing GSCP (2001) *J Bacteriol* 183:4823–4838
- Seshadri G, Myers GSA, Tettelin H, Eisen JA, Heidelberg JF, Dodson RJ, Davidsen TM, DeBoy RT, Fouts DE, Haft DH, Selengut J, Ren QH, Brinkac LM, Madupu R, Kolonay J, Durkin SA, Daugherty SC, Shetty J, Shvartsbeyn A, Gebregeorgis E, Geer K, Tsegaye G, Malek J, Ayodeji B, Shatsman S, McLeod MP, Smajs D, Howell JK, Pal S, Amin A, Vashisth P, McNeill TZ, Xiang Q, Sodergren E, Baca E, Weinstock GM, Norris SJ, Fraser CM, Paulsen IT (2004) *Proc Natl Acad Sci USA* 101:5646–5651
- Schwabe JW, Klug A (1994) *Nat Struct Biol* 1:345–349
- Meyer J, Moulis J (2001) *Handbook of Metalloproteins*. Wiley, Chichester